Founder effects determine the genetic structure of the water flea Daphnia in Ethiopian reservoirs

Haileselasie, Tsegazeabe H.; Mergeay, Joachim ; Vanoverbeke, Joost; Orsini, Luisa; De Meester, Luc

DOI:
doi.org/10.1002/lno.10678

Founder effects determine the genetic structure of the water flea *Daphnia* in Ethiopian reservoirs

Tsegazeabe H. Haileselasie*[1, 2], Joachim Mergeay[1, 3], Joost Vanoverbeke[1], Luisa Orsini[4], Luc De Meester[1]

[1]Laboratory of Aquatic Ecology, Evolution and Conservation, KU Leuven, Ch. Deberiotstraat 32, B-3000 Leuven, Belgium

[2] Department of Biology, College of Natural and Computational Science (CNCS), Mekelle University, P.O. Box: 231, Mekelle, Ethiopia

[3]Research Institute for Nature and Forest, Gaverstraat 4, B-9500 Geraardsbergen, Belgium

[4]School of Biosciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

*Corresponding author: hilinatsegazeabe@yahoo.co.uk, Fax: +3216324575

Running headline: Founder effects in zooplankton populations

Abstract

Founder effects introduce stochasticity in the genetic structure of species at the regional scale. To the extent that founder effects are important that they will result in a reduced signature of space, time and environmental variation in landscape genetic data. We studied the metapopulation genetic structure of recently founded populations of the microcrustacean *Daphnia sinensis* in ten Ethiopian water reservoirs. We used three different approaches of estimating the number of effective founders applied to two independent genetic marker sets to investigate the role of founder effects and to estimate effective size of the founding population. Estimates of founding sizes rarely exceeded eight individuals but were most often limited to less than four individuals. No associations of genetic identities, gene frequencies, measures of genetic diversity or differentiation with environmental and spatial variables were found. Age and size of the reservoirs were not correlated with genetic diversity measures or number of founders in these reservoirs. These findings indicate that neither strong selection, nor dispersal limitation are responsible for the observed pattern of genetic variation. Our results suggest a regional population structure that is strongly impacted by founder events, reflecting colonization by just a few founders per water body, and not noticeably influenced by subsequent dispersal and gene flow. Our results show that rapid colonization of empty habitats and fast population growth by a handful of founders can result in strong founder effects, even in relatively large habitats (estimated populations sizes of several million individuals) that are likely regularly reached by new immigrants.

Keywords: colonization, *Daphnia sinensis*, effective population size, founder effects, metapopulation, monopolization, zooplankton

Introduction

Metapopulation theory describes the interplay between colonization and extinction rates on patch occupancy as a function of death, birth and dispersal rates (Hanski 1998; Levins 1969). Population genetics, on the other hand considers occupied patches, and considers how gene flow and population size interact to influence genetic structure within and among demes of a metapopulation (Wright 1951). In reality, both extinction-colonization dynamics of local populations and changes in genetic diversity by gene flow and drift within these local populations act simultaneously. Dispersal, which is one of the most fundamental processes in ecology, affects many aspects of evolution and population genetics if translated into successful colonization (Bilton *et al.* 2001). Dispersal allows individuals to establish new populations in an empty patch and promotes range expansion following colonization of new sites.

In a metapopulation genetics context, the colonization of an empty habitat patch by founders can be considered as a special case of gene flow. As the new local population grows in size to carrying capacity, new neutral immigrants (having equal expected fitness) can still enter the population, but their relative contribution to the local gene pool is expected to decrease as local population size at the time of immigration becomes larger. In this initial colonization scenario at least two processes are responsible for successful colonization and establishment of a population in an empty patch. First, the response of the immigrants to the local environmental conditions of the habitat they colonize. Second, differences in their time of arrival at the site, which generates a numerical advantage to the first colonizers over late-comers (Boileau *et al.* 1992; De Meester *et al.* 2002). In contrast to the relative ease of establishment of founders, which experience no to a little competition, individuals attempting to immigrate into an established population close to carrying capacity are faced with strong intraspecific competition by the resident population and low levels of resources. Due to this, realized rates of gene flow may be

much lower than expected based on the rates of dispersal (De Meester *et al.* 2002). This reduced establishment success may strongly contribute to prolonged persistence of founder effects (Boileau *et al.* 1992; De Meester *et al.* 2002; Ventura *et al.* 2014). Because founder effects represent a type of sampling error, they introduce stochasticity in the genetic structure at the regional scale, which tends to result in a reduced signature of space and environmental variation in the genetic data (Orsini *et al.* 2013).

Estimates of the number of founders represent baseline estimates for ecological dispersal rates (m_c, the observed number of migrants), which represent the maximal potential for gene flow (m_e, the effective number of migrants) among populations. Gene flow, the realized effect of ecological dispersal on genetic structure, can be estimated indirectly through population genetics as well (Broquet & Petit 2009). Although we know from many population genetic studies in zooplankton that gene flow (m_e) is often much lower than expected, we have relatively few good estimates of ecological dispersal rates (m_c), because they are so hard to measure directly (Bilton *et al.* 2001). Nevertheless, good dispersal estimates provide baseline information for a broad array of ecological and evolutionary studies (Broquet & Petit 2009). Distinguishing between dispersal and gene flow is essential, especially for biological conservation of populations and species.

Here we take advantage of the recent creation of water reservoirs in Northern Ethiopia and the colonization of these water bodies by zooplankton to estimate the number of founders of populations of a zooplankton species using genetic methods. The reservoirs studied here are young (6-18 years) and two to three orders of magnitude larger than most other systems studied so far on founder effects in zooplankton and small invertebrates (Boileau *et al.* 1992; Haag *et al.* 2006; Louette *et al.* 2007). Specifically, we present patterns of genetic composition and differentiation of the water flea *Daphnia sinensis* in reservoirs that range in size from 1.8 to 45.4

hectare. Using variation at nuclear (nDNA) and mitochondrial (mtDNA) genetic markers, we estimate allele frequencies, within-population genetic variation and among-population genetic differentiation, and relate genetic variation and genotype composition to spatial, environmental and temporal variables. We use various methods to independently estimate founding population sizes, and thereby provide baseline estimates of dispersal rates. Using information on the observed genetic structure (F_{ST}) and the associated expected gene flow at various levels of migration-drift equilibrium and ages of populations, we show that the dispersal rates (m_c) are orders of magnitude higher than the actual gene flow rates (m_e).

Methods

Study region and sampling

The studied reservoirs are part of a set of reservoirs constructed between 1984 and 2001 in the highlands of Tigray Regional State, Northern Ethiopia. The rainfall in Tigray region is seasonal and erratic resulting in moisture stress that hampers the rain-fed agriculture (Haregeweyn *et al.* 2006). To solve this problem agricultural development through irrigation has been a priority for the Regional Government of Tigray. Hence, the target of the construction of reservoirs was mainly to bring food self-sufficiency to the area through irrigation but also to use the water for household consumptions (Asmelash *et al.* 2007). Thirty-two of these reservoirs have been the subject of a detailed limnological survey (Asmelash *et al.* 2007; Dejenie *et al.* 2008). Apart from a single natural lake, not inhabited by the focal species of this study, *Daphnia sinensis*, no similar large and deep aquatic systems are known from Tigray (Dejenie *et al.* 2008). Naturally, this species occurs in temporary pools and ponds as well as larger temporary ponds and lakes (Gu *et al.* 2013). The rapid colonization of these reservoirs shortly after their creation by a considerable number of zooplankton taxa, including typical lake species (Dejenie et al. 2008), despite a regional lack of similar habitats suggests that dispersal rates are relatively high and long-distance dispersal events are rather frequent. Water birds (members of the Podicipidae, Pelecanidae, Ciconiidae, Anatidae and Charadriidae family) are common in and alongside the reservoirs (Asmelash *et al.* 2007) and are probably important vectors of dormant propagules of zooplankton (Figuerola & Green 2002).

Thirty-two reservoirs were sampled for zooplankton in September 2005 (Dejenie et al. 2008). Ten of these samples contained *D. sinensis* in large enough numbers for population genetic analyses (see Table S1, Supporting information). In addition, five temporary natural wetlands were sampled, two of which contained *D. sinensis* (henceforth called T1 and T3),

bringing the total number of independent samples to twelve. All *Daphnia* samples were preserved in 100% ethanol until further processing. Although previously identified as *D. carinata* King by Dejenie *et al*. (2008), DNA barcoding indicates that individuals from these reservoirs belong to *D. sinensis* , a member of the *Daphnia similis* species complex (Popova *et al.* 2016). Measurements of geographic position and morphometric, physical, chemical, and biotic variables were recorded for each sampled reservoir (see Table S1, Supporting information; Dejenie *et al*. 2008). Age of the reservoirs was expressed as number of years at sampling time since construction of the reservoir. The two natural populations T1 and T3 were first excluded from all age related analysis, and were in a second analysis arbitrarily given the same age as the oldest reservoir.

Genotyping

DNA of individual *Daphnia* was extracted using the HotShot protocol (Montero-Pau *et al.* 2008) Sample sizes ranged from 6 to 35 individuals per population for mtDNA (Table 1; 10 out of 12 samples with 17 or more individuals) and from 15 to 34 individuals per population for nDNA (Table 2). Differences in sample sizes between both markers are due to unsuccessful amplification with either approach. A fragment of 341 nucleotides of the mtDNA cytochrome oxidase gene, subunit 1 (*COI*) was amplified using primers *SCox1F1* (GGC CCC AGA TAT GGC TTT) and *SCox1R2* (GCT CCA GCT AAT ACT GGT AAA CTT), specifically designed for this study. The polymerase chain reaction mix of 25 µl contained 2 µl DNA, 2.75µl 10x PCR buffer (10 mM Tris-HCl; pH 8.3; 50 mM KCl), 0.4 µM of each primer, 2.2 mM $MgCl_2$, 0.2 mM of each dNTP, and 1 unit Silverstar *Taq* DNA polymerase (Eurogentec®, Liege Belgium). PCR cycling conditions, PCR product purification and sequencing followed the methods of Mergeay *et al*. (2007). The purified fragments were sequenced using 3.2 pmol of *SCox1F1* primer and the

ABI Big Dye Terminator Kit. Sequences were aligned and trimmed in Mega 4.1 (Kumar *et al.* 2008).

Variation at six microsatellite loci using primers originally developed for the related species *Daphnia magna* (*B088*, *B172*, *B087*, *S6-38*, *B064* and *Dma15* (Agostini *et al.* 2010; Jansen B *et al.* 2011) was assessed in a single multiplex PCR reaction of 10 µl consisting of 5 µl HotStar *Taq* DNA polymerase buffer (Qiagen®, Hilden Germany), 0.15 µM, 0.5 µM, 0.3 µM, 0.2 µM, 0.1 µM and 0.3 µM of each primer of locus *Dma15*, *B087*, *B064*, *S6-38*, *B088* and *B172*, respectively, and 2 µl of template DNA. Cycling conditions were 15' hot start denaturation at 95°C followed by 30 cycles of 30" for each step at 95°C, 56°C and 72°C, and a final elongation step at 60°C for 30'. Polymorphism was assessed on an ABI PRISM® 3130 genetic analyser (Applied Biosystems®, Foster City, CA, USA), using an internal Liz Gene-scan size standard by means of the Genemapper 4.0 software (Applied Biosystems®, Foster City, CA, USA).

Population genetic data analysis

Nucleotide diversity of the sequenced *COI* fragments was calculated in DNAsp version 4.5 (Rozas *et al.* 2003). Because we were not interested in the evolutionary relationships among haplotypes that originated thousands of years prior to the colonization of these reservoirs, no attempts were made to construct haplotype networks or to calculate genetic differentiation among populations based on haplotype identity. The observed haplotype frequencies were primarily used to estimate the number of founders involved in the colonization of each reservoir. We calculated observed haplotype richness (HR) for each sample as well as haplotype diversity (HD). HD was calculated as the true diversity equivalent of the Simpson concentration (Jost 2007). These values were compared to expected HR and HD under the null hypothesis that all reservoirs form one panmictic population, using 10^4 permutations in Partition (Veech & Crist 2009). This yields alpha

(local gene diversity) and beta (average differentiation) estimators that are converted to their true diversity equivalents (Jost 2007).

Standard measures of genetic diversity (number of alleles, allelic richness per locus and per population across all loci, observed heterozygosity and expected heterozygosity) at six microsatellite loci were assessed in R using diveRsity package (Keenan *et al.* 2013). Identical multilocus genotypes, on the basis of the combined information of six microsatellite loci, in a given water body were considered to belong to a single clone. Clonal diversity (CD) was expressed as the true diversity equivalent of the Simpson concentration, clonal richness (CR) as the number of multilocus genotypes per water body. Moreover, relative clonal richness was calculated per sample corrected for sample size expressed as proportion of clones to total individuals genotyped as R = (G-1) / (N-1), where G is the number of genotypes and N indicates sample size. We used HWclon (De Meester & Vanoverbeke 1999) to estimate whether or not observed levels of CD and CR were significantly different from a random distribution, given the genetic diversity and allelic polymorphism in the population (De Meester & Vanoverbeke 1999; Vanoverbeke & De Meester 1997).

The standardized genetic variance among populations (F_{ST}) was calculated according to Weir & Cockerham(1984). We used 500 bootstrap pseudoreplicates to estimate 95% confidence intervals of the F_{ST} values. Genetic structure was assessed using the unbiased estimators of Nei & Chesser (1983) of the overall gene diversity (H_T) and subpopulation gene diversity (H_S).

Spatial, environmental and temporal correlates of genetic differentiation

To investigate the role of spatial and environmental variables separately and to disentangle the unique contribution of each variable matrix to the genetic structure of the studied populations we used redundancy analyses (RDA, a linear constrained ordination technique Dray

et al. 2006). In a multivariate variation partitioning analysis, the contributions of local environmental predictors (n= 16 variables provided in Table S1, Supporting information) and spatial predictors are tested by generating adjusted redundancy statistics (R^2_{adj}) in an RDA analysis. A significant effect of environment would imply sorting of clones with different traits and niches along environmental gradients.

Under a model of persistent founder effects, we expect genetic structure to be mainly caused by chance events as dispersal is likely not limiting at the investigated spatial scale. Hence, we expect to find at most a weak spatial genetic structure in the data (Orsini *et al.* 2013). To test this, with the nDNA allele frequency data we performed a principle coordinates analysis (PCoA) and then used the population loadings of the first six PCoA axes as dependent variables in a distance-based redundancy analysis (db-RDA). In this RDA we attempted to explain the observed genetic variation as a function of distance-based eigenvector maps (dbMEM) (Dray *et al.* 2006). This analysis allows to find spatial patterns in the genetic data other than linear ones, making this a more powerful approach with lower type II error rates than Mantel tests (Legendre & Fortin 2010). Five positive dbMEM eigenvectors were retained and were used as explanatory variables in a forward selection procedure. Although this particular approach has the risk of identifying a false positive spatial signal (see Blanchet *et al.* 2008), the double stop criterion of Blanchet *et al.* (2008) is very conservative with regard to small datasets. Here we take a more liberal approach, involving forward selection of spatial variables without prior testing of the overall spatial model, to make sure that any lack of a detectable spatial signal is not due to the use of conservative statistical methods.

In parallel, we performed a separate RDA relating the mtDNA data to the spatial data (dbMEM). These mtDNA data were Hellinger-transformed to allow the use of linear regression analyses in zero-inflated data (Legendre & Gallagher 2001).

Using non-parametric correlation (Spearman's rho) analyses, we related diversity measures (H_e, AR, HR, HD, CR, CD, number of founders at mtDNA, and number of founders at nDNA) to age, depth and size of the reservoirs. Under a model of persistent founder effects, genetic diversity should not be related to age or size of the reservoir, as late-arriving immigrants are expected to have little impact on the genetic structure compared to the very first founders. If founder effects still persisted at the time of sampling, we expect that old and large populations do not differ in genetic structure from younger and/or smaller populations. Specifically, we tested if a model of population structure including reservoir age, reservoir size or their interaction explained the genetic structure better than the null model not including age or size using Geste v. 2 (Foll & Gaggiotti 2006). Geste calculates a specific F_{ST} value for each population that represents the population specific contribution to the total genetic differentiation in the metapopulation. Then, the effect of age and/or size on these F_{ST} values is evaluated using generalized linear models. The posterior probability of each model was used to select the model with the highest probability given our data.

Estimating the number of effective founders

We used three different approaches to estimate the number of effective founders, which reflects how many individuals contributed to the observed genetic diversity in each local population. First, we used a general approach based on F-statistics from microsatellite data, using the principle that the inbreeding coefficient among populations just after colonization is $F_{ST} = (2K)^{-1}$, with K the average number of founders per population (Boileau *et al.* 1992; Wade & McCauley 1988). Confidence intervals (95% CI) for F_{ST} were calculated by bootstrapping over loci (500 replicates). This method provides the average effective size of the founding population.

Second, we used a general simulation approach in the programming environment R (R 2.14; The R Foundation for Statistical Computing, 2014) to calculate the expected HR and HD under a model of random colonization from a regional gene pool with 1, 2... 10, 15 and 20 founders. Expected HR and HD values were calculated by randomly sampling 10^4 times the pre-set number of founders in 200 populations from an estimated regional frequency distribution, which was based on the actual haplotype counts over all water bodies, or on presence-absence data for each haplotype per water body. For each of the random samples we calculated the probability that the expected average HR or HD, over the 200 population, was smaller or larger than the values observed in our empirical dataset. The product of these values provides the overall probability that the observed HD or HR is achieved by the corresponding number of founders. The R-script is available as supplementary information (Table S2, Supporting information), and provides the average census size of the founding population. Third, we used a population-specific approach for which we used the Colonize script (Mergeay *et al.* 2007; Vanoverbeke & Mergeay 2007). This is a standalone command-line tool that calculates the likelihood that a predefined number of founders from a predefined source population established the focal population, given the gene frequency distribution of the source and the sink populations. The number of founders associated to the highest likelihood score provides the best estimate for the founding propagule size of that population. Again, this provides the census size of the founding population, which ignores that some individuals contributed less to the genetic structure of the founding population than other individuals. For each water body, we calculated likelihood scores for one to thirty founders, and set both the number of batches and the number of random samples to 500. Ideally, one has multiple putative source populations from which to sample, so as to assign the most likely source population as well (see Mergeay et al. 2007, for an example). Here we have no such prior information, and hence we use the regional gene pool (the average over all our samples) as the

overall source population. To increase the overall robustness of this approach, three different prior allele frequency distributions were used. 1) Using the regional frequency of each allele over the pooled data of all investigated water bodies (distribution = Freq.). Here we used the rare allele correction in Colonize to account for extremely rare alleles. 2) Using presence-absence of each allele per population and counting the frequency of occurrence of each allele over all populations (distribution = Rich.). This approach gives less weight to alleles that dominate in certain populations but are rare in other populations. 3) Using three abundance classes for the regional allele frequencies (1: frequency <15%; 2: frequency 15-30%; 3: frequency >30%). This approach (distribution = Level) gives even more weight to rare alleles. Analyses with Colonize were performed separately for the mtDNA data and for the nDNA data in order to obtain independent estimates for both marker types. Overall, these three approaches yield one overall estimate based on F_{ST} (first method), two overall estimates based on mtDNA haplotype richness and haplotype distributions (second method), and per reservoir three estimates based on mtDNA and three based on nDNA (third method).

Testing assumptions

All of the outlined methods to estimate the number of founders rely on similar assumptions, but vary in their sensitivity to violations thereof. Here we outline how we tested for violations of the assumptions. First, we assume that genetic drift has not yet strongly affected allele frequency distributions (especially fixation or loss of alleles), given that they were founded at most 6 to 18 years before sampling. Second, we assume that all founders are genetically independent.

The main source of genetic drift in cyclical parthenogens like *Daphnia* is clonal selection

289 (reduction in clonal and genetic diversity as a result of selection among clones in the population)
290 (De Meester *et al.* 2006), which may erode genetic diversity considerably and thereby reduce our
291 estimates of founding population sizes. To assess whether clonal population structure (leading to
292 a similar signal as genetic bottlenecks) has affected our results, we performed a Spearman Rank
293 order correlation between clonal diversity (CD) and clonal richness (CR) versus the estimated
294 number of founders per population. This was performed for all estimates of the number of
295 founders, which have different sensitivities for common and rare alleles. Significant correlations
296 would reflect that clonal erosion affects our estimates of founding population sizes. Furthermore,
297 because drift reduces richness faster than diversity (Cornuet & Luikart 1996), founding
298 population size estimates based on richness should be lower than those based on diversity indices
299 if genetic drift is really important. Next, genetic drift affects mitochondrial genetic structure
300 stronger than nuclear genetic structure, because the effective population size at mitochondrial
301 genes is smaller (Hamilton 2011). Hence our founding number estimates should be lower when
302 using mitochondrial data if these were strongly influenced by genetic drift. We used the Student
303 t-test to test for differences among all these cases.

Results

Genetic diversity

Among a total of 285 sequences we found 25 polymorphic positions out of 299 nucleotide positions in the *CO1* gene fragment. This resulted in six distinct mitochondrial haplotypes. Two of these haplotypes, H5 and H6, were singletons detected from Adi Gela and Adi Kenafiz, respectively. Four haplotypes were common, with overall frequency of occurrence of 36.0, 33.0, 20.3 and 9.70 % for H2, H1, H3 and H4, respectively (Table 1 and Fig. 1). Most populations, including both natural systems, were dominated by one or two haplotypes (average HD = 1.74). Overall, the observed haplotype diversity or richness in a given population was always significantly lower (X^2 text, $p < 0.0001$) than the expected diversity or richness assuming a panmictic regional metapopulation (Table 1). Pairwise nucleotide diversity among haplotypes ranged from 0.003 to 0.047 (overall nucleotide diversity = 0.021).

For the microsatellite markers, we found an average of 5.3 ± 3.6 (± standard deviation) alleles per locus over the whole metapopulation, whereas the mean allelic richness per locus was 2.58 ± 0.5 per population. The number of alleles per locus ranged from 2 to 11, with a total of 32 alleles scored over the six microsatellite loci combined. The total number of alleles observed across all loci per population ranged from 11 to 20, with a mean allelic richness of 2.32 alleles per locus (Table 2). The observed heterozygosity for the 12 relatively young *Daphnia* populations ranged from 0.21 to 0.57 whereas the expected heterozygosity (H_e) ranged from 0.24 to 0.54 per population (Table 2 and Table S3, Supporting information).

In total, we found 183 unique multilocus genotypes (MLGs) out of 293 individuals successfully genotyped. The majority of those MLGs (88%) were represented by a single individual whereas 5% of the MLGs (n= 10) were represented by two individuals. Only a small

number of MLGs (n = 18) was shared between reservoirs. The highest clonal richness (CR = 27) and clonal diversity (CD = 22.3) was observed for Gum Selasa (Table 2). The difference between the observed clonal richness/diversity and expected clonal richness/diversity was not statistically different at $\alpha = 0.05$ for all the 12 populations studied, indicating that there is no substantial clonal erosion (Table 2).

All values of pairwise genetic differentiation (F_{ST}) were significant ($p < 0.05$). Nearby population pairs were not more related to each other than distant pairs (Table S4, Supporting information). The highest pairwise F_{ST} value was observed in the comparison between T1 and Adi Kenafiz ($F_{ST} = 0.585$), while the lowest pairwise F_{ST} value ($F_{ST} = 0.037$) was between Gereb Awso and Dibla (Table S4, Supporting information).

None of the RDA analyses yielded a model with one or more spatial or environmental explanatory variables that could significantly ($p < 0.05$) explain the variation in the genetic data, either in the distribution of the mtDNA haplotypes, or in the allele frequency data of the microsatellite loci. Exclusion of the two natural systems did not affect the general pattern. Mantel tests between pairwise genetic distance (Nei's genetic distance) and geographic distances or environmental distances yielded correlation coefficients of $r = -0.114$ ($p = 0.658$) and $r = 0.181$ ($p = 0.212$), respectively, thus confirming the absence of any spatial trend in the genetic data (Fig. 2).

Number of founders

Method 1: F_{ST}-based. The overall among-population fixation index (F_{ST}) was 0.237, with 95% confidence intervals (CI) ranging from $0.180 < F_{ST} < 0.342$. Without T1 and T3, F_{ST} equalled 0.219 (95% CI: $0.169 < F_{ST} < 0.309$). Since $F_{ST} \approx 1/2N$ at colonization, this reflects average effective founding population sizes of 2.3 individuals (95% CI: $1.6 < Ne < 3.0$). Put differently, the average genetic diversity we observed corresponds to a mean effective founding population size of 1.6 to 3.0 individuals.

Method 2: Comparing observed to expected richness and diversity estimates. We found an average observed haplotype richness HR = 2.5 and an average observed haplotype diversity HD = 1.74 (Table 1). When comparing the average observed levels of HR and HD to expected HR and HD, we found that average founding population size estimates smaller than two and larger than eight are improbable at p-value = 0.05 (Table S5, Supporting information). The highest probability scores were obtained with 4, 3, 3 and 4 founders for the four types of simulations (Table S5, Supporting information).

Method 3: Population-specific simulations. We used population-specific simulations using mtDNA and nDNA, based on three prior theoretical allele frequency distributions (Freq, Rich, and Level, in descending order of sensitivity to rare alleles). For mtDNA, all three prior allele frequency distributions yielded very comparable estimates (Table 3 and Table S6, Supporting information), with averages ranging between 3.5 and 4 founders ($1 \leq \text{range} \leq 8$). This didn't change appreciably when the natural systems were excluded (results not shown). For nDNA, similar average (2.8 to 5.5) values were found ($2 \leq \text{range} \leq 13$), although the estimate using the Freq. prior distribution was somewhat higher and was positively skewed due to higher estimates

for two populations (Table 3 and Table S6, Supporting information). Confidence in the prior distribution of regional allele frequencies (Freq.) was unacceptably low (highest observed likelihood score < 0.05) in six cases for nDNA and two cases for mtDNA. The prior distribution based on local richness of alleles (Rich) yielded one estimate at nDNA with too low likelihood scores (Table 3 and Table S6, Supporting information). All these estimates indicate that founding population sizes were typically smaller than five individuals, and very rarely exceeded ten individuals.

Effect of size and age of the reservoir

There was no significant relation ($p > 0.05$) between H_e and surface area or log (surface area) of the water body ($S = 183.28$, Spearman rank $r = 0.3563$, p-value = 0.126) or with age of the reservoir ($S = 342.35$, Spearman rank $r = -0.197$, p-value = 0.730, Figs S1, Supporting information). The only significant correlation found in a total of 42 associations tested (age, depth, area, versus AR, H_e, H_o, HR, HD, CR/N, CD/N, F_{IS}, number of founders at mtDNA (Colonize-Rich), and number of founders at nDNA (Colonize-Rich)) was between H_e and average depth ($r = -0.79$, $p < 0.001$). However, after Bonferroni correction, this p-value was larger than 0.05. All other correlations were extremely weak and statistically insignificant at $\alpha=0.05$ (absolute value of $r < 0.15$, uncorrected $p > 0.10$; Figs S1, Supporting information). Inclusion of size and/or age of the reservoir did not provide a better model (Geste v. 2) for genetic structure than the more parsimonious null model. All of these results support the hypothesis that founder effects are the main drivers of genetic structure and indicate that clonal genetic drift did not markedly influence our estimates of founding population sizes. In addition, none of the tests we did could show a significant difference between founding population size estimates based on mitochondrial versus nuclear genetic data, or based on richness versus

diversity estimates (all p values > 0.05). Thus, we have no indication that the assumptions concerning genetic drift were violated (see Table S7 and Fig. S1, supporting information).

Discussion

Although we included a broad set of environmental and spatial variables, we found no pattern with environmental variation, space and time (age) in the distribution of genetic variation of the studied *Daphnia sinensis* populations inhabiting reservoirs in Tigray. Both the nuclear and mitochondrial markers that we used are expected to behave neutrally. As such, we expected a stronger signature of space than of environment. Still, a correlation with environmental variables may result when particular haplotypes would hitchhike with particular genotypes or fixed allele combinations that are favoured under certain conditions. Two studies on strictly asexual zooplankton with comparable sample sizes and statistical power found clear environmental and/or spatial structuring in their studies (Aguilera *et al.* 2007; Pantel *et al.* 2011) suggesting that the lack of patterns in our dataset is not merely a consequence of insufficient statistical power. Evolution-mediated priority effects, the key feature of the monopolisation hypothesis (De Meester *et al.* 2002; De Meester *et al.* 2016), are expected to be less important in asexual taxa than in similar sexual taxa, due to a reduced ability for rapid local adaptation in the asexuals. As a consequence, it is expected that fitness differences among dispersed clones will lead to a match between environmental gradients and landscape genetic structure in asexual taxa, similar to species sorting in communities (De Bie *et al.* 2012; Leibold *et al.* 2004). Conversely, if the colonizing propagules of a sexual species harbours sufficient genetic variation to allow local genetic adaptation, the increased fitness of resident populations may reduce establishment success of new immigrants and thus reduce gene flow (De Meester *et al.* 2016). As a result, the match between environmental and genetic variation is expected to be less strong in sexual than in asexual species, which is the emergent pattern from our study on a cyclically parthenogenetic *Daphnia* species and contrasts with the two aforementioned studies that focused on obligately parthenogenetic *Daphnia*.

To investigate the number of founders typically involved in the colonization of new moderately-sized freshwater systems (ranging from 1.8-45.4 ha in size), we used three different approaches that rely on different test statistics with varying prior parameters, applied on two independent sets of genetic markers. All approaches indicated that typically less than five founders per habitat were responsible for the observed pattern of genetic diversity in the studied reservoirs. Irrespective of whether we used estimates based on richness data or more detailed frequency data, we obtained very similar estimates, showing that our results are robust to strong allele frequency changes that may have occurred since colonization. Admittedly, the different approaches we used all rely on similar assumptions, including an absence of genetic drift since colonization, and genetic independence of each founder. The second assumption that the founders are genetically independent from each other may have been violated to some extent. Birds, for example, may disperse more than one dormant stage at the same time from a single source, thereby introducing multiple related propagules. Especially results from mtDNA are expected to be prone to such bias, given the much lower local and regional genetic variation compared to the levels of variation found at nDNA. Estimated numbers of founders for nDNA and mtDNA were, however, very similar.

We have detected high genetic differentiation among population (F_{ST}= 0.232) and no isolation by dispersal limitation. This indicates low levels of gene flow among populations. Furthermore, we failed to detect isolation-by-environment (IBE), which rules out the possibility that sorting of genotypes along environmental gradients similar to species sorting in communities (Leibold *et al.* 2004) might have driven the observed high genetic differentiation among populations. Thus, our results support the idea that colonization dynamics in a newly created metapopulation are strongly affected by founder effects exerted by a limited number of founding genotypes. The founder effects observed here indicate that metapopulation and colonization

446 dynamics in this species resemble a lottery model (Sale 1977). In Sale's (1977) lottery model,
447 individuals compete for a limited number of discrete resources and once a resource is claimed, an
448 individual cannot be usurped from it. The classic lottery model was formulated at the community
449 level and with respect to microsites. However, it here acts at the level of genetic variants of a
450 species, and at the habitat level in a metapopulation. Populations are thus founded by a small
451 number of individuals from a varied array of regional sources. As long as a local genetic variant
452 persists (also if persistence is mediated through dormant stages; (Mergeay *et al.* 2007), the niche
453 space will continue to be occupied by these local variants, thereby pre-empting niche space for
454 immigrants. Several empirical studies focusing on colonization of novel habitats have shown that
455 dispersal rates in zooplankton are high (Cáceres & Soluk 2002; Jenkins & Buikema 1998;
456 Louette & De Meester 2005). The lack of spatial genetic patterns in our dataset also suggests that
457 dispersal per se is not limiting at the spatial scale here studied.

458 The sole environmental variable that showed a significant but negative correlation with
459 genetic diversity was average lake depth. One may speculate that the negative correlation
460 between depth and H_e reflects a species-specific preference for shallow waters, thereby reducing
461 the likelihood that a colonizing propagule will survive in deep reservoirs. This is indeed expected
462 from an organism that seems to naturally inhabit shallow pools. In that case, however, we would
463 also expect a similar negative relation between depth and number of founders, or other measures
464 of genetic diversity, which was not the case. An alternative explanation is that deeper lakes result
465 in more stable habitat conditions and therefore in populations that survive year-round and are
466 thus less dependent on dormant egg banks for survival. It is well known that more permanent
467 populations in *Daphnia* exhibit lower genetic diversity because of ongoing clonal erosion (De
468 Meester *et al.* 2006; Hebert 1987).

469 Earlier studies (Boileau *et al.* 1992; Haag *et al.* 2006) already showed that founder events

can strongly determine metapopulation structure, but the habitats they studied were very small (less than <100 m^2). The systems we study are thousand times larger than the typical size of the small habitats studied earlier, with associated differences in carrying capacity, effective population size, genetic drift and inbreeding. Although the results shown here should be interpreted with some caution given that the limited number of reservoirs that was inhabited by the studied *Daphnia* species resulted in a reduced statistical power in detecting spatial and environmental patterns, our analyses strongly indicate that zooplankton populations of these new large water bodies are typically founded by just a handful of individuals. Interestingly, the number of founders in these reservoirs (on average 4-6) is strikingly similar to the range found in ponds with population sizes that are up to a thousand times smaller (Boileau *et al.* 1992; Louette *et al.* 2007). Similarly, the local recolonization by *Daphnia barbata* of the 150 km^2 large Kenyan Lake Naivasha happened most likely by no more than nine individuals from an old dormant egg bank (Mergeay *et al.* 2007).

Inbreeding effective population size (N_e) in populations is a function of the number of founders and is thus generally small in our zooplankton population. It seems that in zooplankton, habitat size per se, at least within given boundaries, may have little influence on the effective population size. Next to the low number of founders that seem typically involved, our results indicate that these founder effects were equally high irrespective of the age of the reservoirs. Several case studies on the propagule banks of *Daphnia* populations have demonstrated high local genetic stability over periods of 50-150 years (Decaestecker *et al.* 2007; Mergeay *et al.* 2007). Recently, Ventura *et al.* (2014) even provided empirical evidence for founder effects lasting thousands of years. All this evidence indicates that zooplankton populations primarily have founder-controlled populations (Okamura & Freeland 2002), similar to founder-controlled communities (Sale 1977). In such populations, dispersal contributes little to gene flow and is

mostly prevalent during the initial phase of colonization of empty or newly created habitats. While dormant propagules are the main unit of dispersal in most zooplankton, their most pervasive impact on landscape genetic structure may be their role in the short-term and long-term local persistence of populations as well as in fostering colonization of empty habitats rather than that they contribute to continuous gene flow among populations. Even seemingly extinct populations may still be recolonized by local dormant egg banks once the habitat becomes suitable again after decades (Mergeay *et al.* 2007). This has profound consequences for our view on metapopulation biology of zooplankton and other micro-organisms, as these species often share the lack of landscape genetic structure reflecting strong isolation-by-distance (Okamura & Freeland 2002). More specifically, we should not equal high potential for dispersal into high rates of gene flow (De Meester *et al.* 2016). In very small water bodies, however, negative effects of genetic drift and inbreeding can be pronounced, and the positive influence on fitness of immigrant alleles or genotypes from immigrants may then promote immigration and gene flow (Ebert *et al.* 2002). One might therefore expect a shift from a gene flow dominated system in extremely small populations (Ebert et al 2002) to metapopulations that are more strongly dominated by local processes combined with extinction-recolonization dynamics in somewhat larger systems such as the reservoirs studied here, shallow lakes and the sometimes much smaller (approx. 100 m^2) farmland ponds (De Meester *et al.* 2002; Louette *et al.* 2007; Vanoverbeke & De Meester 1997).

Acknowledgments

This study was financially supported by the KU Leuven Research Fund Centre of excellence funding PF/2010/07. Samples were obtained thanks to financial support from the VLIR inter-institutional cooperation (IUC) between Mekelle University (MU) and KU Leuven. The authors

thank T. Dejenie, S. Risch, S. Declerck and others for help with sampling, and B. Hellemans, R. De Schutter and S. Geldof for help with lab work. Two anonymous reviewers provided constructive and helpful comments on an earlier version. JM was supported by a postdoctoral fellowship of the Research Foundation Flanders (FWO).

Data Accessibility

Sampling locations, raw environmental data for each reservoir and microsatellite genotype data is stored in in Dryad®.

References

Agostini C, Agudelo PA, Ba K*, et al.* (2010) Permanent Genetic Resources added to Molecular Ecology Resources Database 1 October 2010–30 November 2010. *Molecular Ecology Resources* **11**, 418-421.

Aguilera X, Mergeay J, Wollebrants A, Declerck S, De Meester L (2007) Asexuality and polyploidy in *Daphnia* from the tropical Andes. *Limnology & Oceanography* **52**, 2079-2088.

Asmelash T, Dejenie T, Declerck S*, et al.* (2007) *Ecological atlas of reservoirs in Tigray, Northern Ethiopia. Tigray Livelihood Papers No. 4* VLIR-Mekelle University IUC Programme and Zala-Daget Project, Mekelle.

Bilton DT, Freeland JR, Okamura B (2001) Dispersal in freshwater invertebrates. *Annual review of ecology and systematics* **32**, 159-181.

Blanchet FG, Legendre P, Borcard D (2008) Forward selection of explanatory variables. *Ecology* **89**, 2623-2632.

Boileau MG, Hebert PDN, Schwartz SS (1992) Nonequilibrium gene-frequency divergence - persistent founder effects in natural populations. *Journal of Evolutionary Biology* **5**, 25-39.

Broquet T, Petit EJ (2009) Molecular estimation of dispersal for ecology and population genetics. *Annual review of ecology and systematics* **40**, 193-216

Cáceres CE, Soluk DA (2002) Blowing in the wind: a field test of overland dispersal and colonization by aquatic invertebrates. *Oecologia* **131**, 402-408.

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**, 2001-2014.

De Bie T, De Meester L, Brendonck L*, et al.* (2012) Body size and dispersal mode as key traits determining metacommunity structure of aquatic organisms. *Ecology letters* **15**, 740-747.

De Meester L, Gómez A, Okamura B, Schwenk K (2002) The Monopolization Hypothesis and the dispersal–gene flow paradox in aquatic organisms. *Acta Oecologica* **23**, 121-135.

De Meester L, Vanoverbeke J (1999) An uncoupling of male and sexual egg production leads to reduced inbreeding in the cyclical parthenogen *Daphnia*. *Proceedings of the Royal Society of London B: Biological Sciences* **266**, 2471-2477.

De Meester L, Vanoverbeke J, De Gelas K, Ortells R, Spaak P (2006) Genetic structure of cyclic parthenogenetic zooplankton populations – a conceptual framework. *Archiv für Hydrobiologie* **167**, 217-244.

De Meester L, Vanoverbeke J, Kilsdonk LJ, Urban MC (2016) Evolving Perspectives on Monopolization and Priority Effects. *Trends in Ecology & Evolution* **31**, 136-146.

Decaestecker E, Gaba S, Raeymaekers JAM*, et al.* (2007) Host-parasite 'Red Queen' dynamics archived in pond sediment. *Nature* **450**, 870-874.

Dejenie T, Asmelash T, De Meester L*, et al.* (2008) Limnological and ecological characteristics of tropical highland reservoirs in Tigray, Northern Ethiopia. *Hydrobiologia* **610**, 193-209.

Dray S, Legendre P, Peres-Neto PR (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *ecological modelling* **196**, 483-493.

Ebert D, Haag C, Kirkpatrick M*, et al.* (2002) A selective advantage to immigrant genes in a *Daphnia* metapopulation. *Science* **295**, 485-488.

Figuerola J, Green AJ (2002) Dispersal of aquatic organisms by waterbirds: a review of past research and priorities for future studies. *Freshwater biology* **47**, 483-494.

Foll M, Gaggiotti O (2006) Identifying the environmental factors that determine the genetic structure of populations. *Genetics* **174**, 875-891.

Gu Y, Xu L, Lin Q, Dumont H, Han B (2013) A new subspecies of *Daphnia*: *Daphnia similoides sinensis*. *Ecological Science* **32**, 308-312.

Haag CR, Riek M, Hottinger JW, Pajunen VI, Ebert D (2006) Founder events as determinants of within-island and among-island genetic structure of *Daphnia* metapopulations. *Heredity* **96**, 150-158.

Hamilton M (2011) *Population Genetics* John Wiley & Sons, Chichester UK.

Hanski I (1998) Metapopulation dynamics. *Nature* **396**, 41-49.

Haregeweyn N, Poesen J, Nyssen J*, et al.* (2006) Reservoirs in Tigray (Northern Ethiopia): characteristics and sediment deposition problems. *Land Degradation & Development* **17**, 211-230.

Hebert PDN (1987) Genotypic characteristics of cyclic parthenogens and their obligately asexual

derivatives. In: *The evolution of sex and its consequences* (ed. Stearns SC), pp. 175-195. Birkhäuser Basel, Basel, Switzerland.
Jansen B, Geldof S, De Meester L, Orsini L (2011) Isolation and characterization of microsatellite markers in the waterflea *Daphnia magna*. *Molecular Ecology Resources* **11**, 418-421.
Jenkins DG, Buikema AL (1998) Do similar communities develop in similar sites? A test with zooplankton structure and function. *Ecological Monographs* **68**, 421-443.
Jost L (2007) Partitioning diversity into independent alpha and beta components. *Ecology* **88**, 2427-2439.
Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA (2013) diveRsity: an R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution* **4**, 782-788.
Kumar S, Nei M, Dudley J, Tamura K (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in bioinformatics* **9**, 299-306.
Legendre P, Fortin M-J (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources* **10**, 831-844.
Legendre P, Gallagher ED (2001) Ecologically meaningful transformations for ordination of species data. *Oecologia* **129**, 271-280.
Leibold MA, Holyoak M, Mouquet N*, et al.* (2004) The metacommunity concept: a framework for multi-scale community ecology. *Ecology letters* **7**, 601-613.
Levins R (1969) Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America* **15**, 237-240.
Louette G, De Meester L (2005) High dispersal capacity of cladoceran zooplankton in newly created communities. *Ecology* **86**, 353-359.
Louette G, Vanoverbeke J, Ortells R, De Meester L (2007) The founding mothers: the genetic structure of newly established *Daphnia* populations. *Oikos* **116**, 728-741.
Mergeay J, Vanoverbeke J, Verschuren D, De Meester L (2007) Extinction, recolonisation and dispersal through time in a planktonic crustacean. *Ecology* **88**, 3032-3043.
Montero-Pau J, Gómez A, Muñoz J (2008) Application of an inexpensive and high-throughput genomic DNA extraction method for the molecular ecology of zooplanktonic diapausing eggs. *Limnology and Oceanography: Methods* **6**, 218-222.
Nei M, Chesser RK (1983) Estimation of fixation indices and gene diversities. *Annals of Human Genetics* **47**, 253-259.
Okamura B, Freeland JR (2002) Gene flow and the evolutionary ecology of passively dispersing aquatic invertebrates (J.M. Bullock, R.E. Kenward, R.S. Hails, eds.). Blackwell Science, Oxford. In: *Dispersal Ecology: 42nd Symposium of the British Ecological Society* (eds. Bullock JM, Kenward RE, Hails RS). Blackwell Science, Oxford, UK.
Orsini L, Vanoverbeke J, Swillen I, Mergeay J, De Meester L (2013) Drivers of population genetic differentiation in the wild: isolation-by-dispersal limitation, isolation-by-adaptation and isolation-by-colonization. *Molecular Ecology* **22**, 5983-5999.
Pantel JH, Juenger TE, Leibold MA (2011) Environmental gradients structure *Daphnia pulex* × *pulicaria* clonal distribution. *Journal of Evolutionary Biology* **24**, 723-732.
Popova EV, Petrusek A, Kořínek V*, et al.* (2016) Revision of the Old World *Daphnia* (Ctenodaphnia) *similis* group Cladocera: Daphniidae). *Zootaxa* **4161**, 1-40.
Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496-2497.
Sale PF (1977) Maintenance of high diversity in coral reef fish communities. *American Naturalist* **111**, 337-359.
Vanoverbeke J, De Meester L (1997) Among-populational genetic differentiation in the cyclical parthenogen *Daphnia magna* (Crustacea, Anomopoda) and its relation to geographic distance and clonal diversity. *Hydrobiologia* **360**, 135-142.
Vanoverbeke J, Mergeay J (2007) Colonize: a tool to estimate the number of individuals colonizing a new habitat. *Ecological Archives*, E088-188-S081.
Veech JA, Crist TO (2009) PARTITION: software for hierarchical additive partitioning of species diversity, v. 3.0. Miami University, Oxford, Ohio USA.
Ventura M, Petrusek A, Miró A*, et al.* (2014) Local and regional founder effects in lake zooplankton persist after thousands of years despite high dispersal potential. *Molecular Ecology* **23**, 1014–1027.

635 Wade MJ, McCauley DE (1988) Extinction and recolonization: their effects on the genetic differentiation of
636 local populations. *Evolution* **42**, 995-1005.
637 Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*
638 **38**, 1358-1370.
639 Wright S (1951) The genetical structure of populations. *Annals of Human Genetics* **15**, 323-354.
640
641

642 List of tables

643 Table 1: Genetic diversity in mtDNA haplotypes: observed frequencies (expressed as fractions)
644 of each haplotype per water body and diversity descriptors. N: number of individuals extracted
645 per sampling site. HR: haplotype richness. HD-Si: true haplotype diversity measured with the
646 Simpson index. Average observed alpha diversity is the average observed within-sample
647 diversity weighted by sample size. Average expected alpha diversity gives the expected value of
648 HR or HD given a panmictic population over all water bodies, using 10,000 permutations. The
649 range of the expected values shows the lowest and highest value among all permutations over
650 individuals. All observed values deviate significantly ($p<0.0001$) from expected values. True beta
651 diversity is calculated as gamma/alpha.
652

		Haplotype n°							
Water body	N	1	2	3	4	5	6	HR	HD-Si
Adi Gela	17	0	0.71	0.06	0.18	0.06	0	4	1.86
Adi Kenafiz	18	0	0	0.39	0.56	0	0.06	3	2.16
Dibla	27	0.26	0.07	0.67	0	0	0	3	1.93
Gereb Awso	35	1.00	0	0	0	0	0	1	1.00
Gereb Mihiz	31	0.45	0.16	0.39	0	0	0	3	2.63
Gum Selasa	33	0.27	0.06	0.27	0.39	0	0	4	3.25
Haiba	6	0.83	0	0.17	0	0	0	2	1.38
Mai Leba	29	0.03	0.93	0	0.03	0	0	3	1.15
Meala	32	0.97	0.03	0	0	0	0	2	1.06
Tsinkanet	12	0	1.00	0	0	0	0	1	1.00
Temp 1 (T1)	22	0.14	0.36	0.50	0	0	0	3	2.49
Temp 3 (T3)	23	0	1.00	0	0	0	0	1	1.00
Overall frequency		0.33	0.36	0.203	0.097	0.005	0.005		
Average observed (alpha)		0.38	0.30	0.19	0.12	0.01	0.01	2.5	1.74
Total diversity (gamma)								6	3.43
Average expected alpha								4	3.40
Range expected alpha								3.7-4.2	3.01-3.67
True beta diversity								2.4	2.10

653

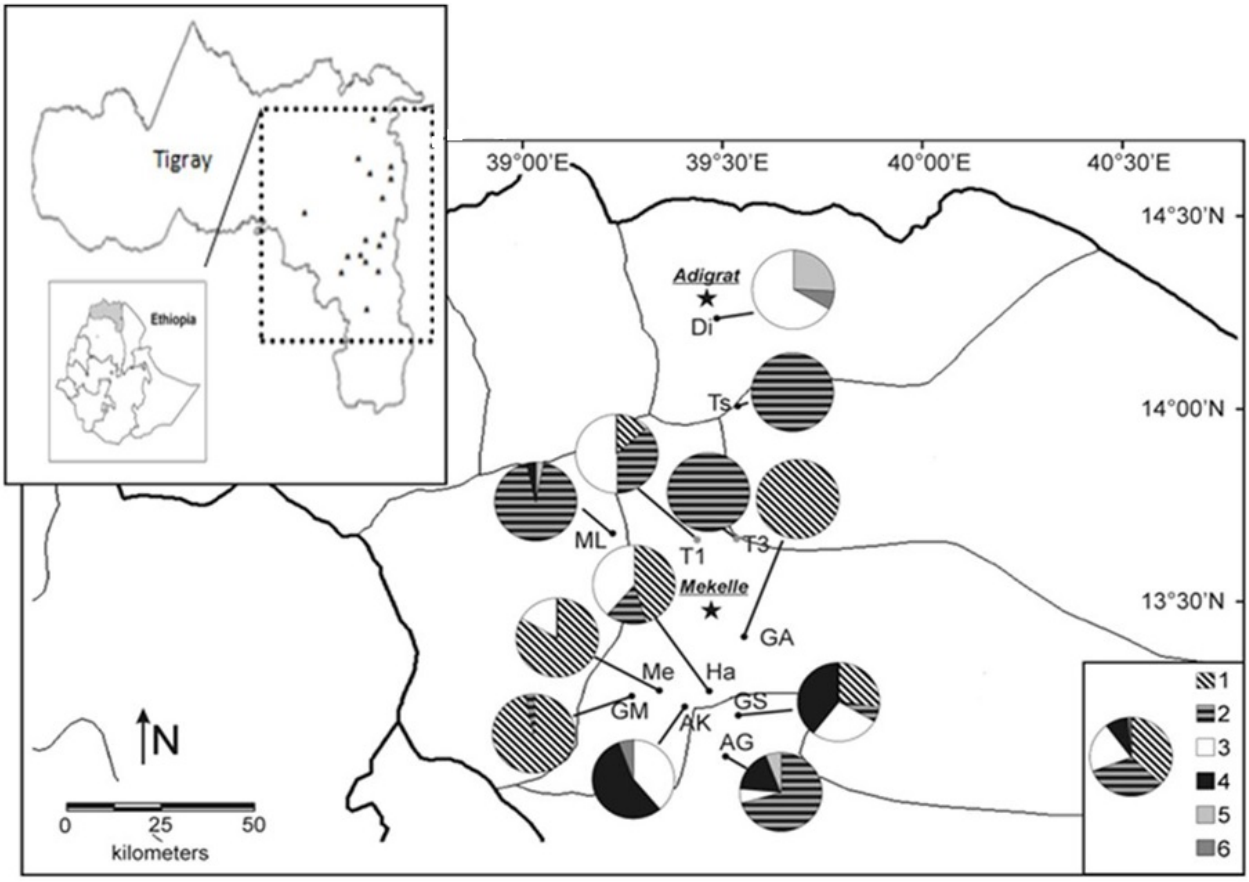
Table 2: Clonal and genetic diversity based on microsatellite loci (nDNA). N: sample size; n: number of individuals with complete genotypic information (6 loci) on which calculations of clonal richness (CR) and clonal diversity (CD) were based. CR=clonal richness; CD=clonal diversity.

Water body	Observed						Expected[$]							
	N	n	CR	CD	CR/n	CD/n	CR ± S.e	CD ± S.e	A	AR	Ho	He	HWE[¥]	F_{IS}
AG	30	20	19	18.18	0.95	0.91	20.92±0.09	20.86±0.02	20	2.97	0.38	0.54	0.001	0.302
AK	30	20	15	11.11	0.75	0.56	12.04±0.03	11.40±0.05	11	1.76	0.39	0.34	0.335	-0.164
DIB	30	27	19	12.79	0.70	0.47	25.13±0.05	22.86±0.08	15	2.29	0.41	0.32	0.085	-0.29
GA	32	31	13	8.50	0.42	0.27	21.35±0.06	16.10±0.08	14	1.95	0.34	0.27	0.108	-0.242
GM	36	34	19	6.64	0.56	0.20	28.09±0.10	25.05±0.08	19	2.46	0.43	0.39	0.001	-0.106
GS	40	32	27	22.26	0.84	0.70	22.93±0.03	22.08±0.05	18	2.63	0.38	0.47	0.000	0.19
HA	16	16	16	16.00	1.00	1.00	14.99±0.01	14.98±0.01	17	2.62	0.35	0.41	0.001	0.145
ML	29	23	22	21.16	0.96	0.92	23.68±0.02	23.42±0.03	13	2.13	0.57	0.47	0.012	-0.19
MA	30	20	3	1.23	0.15	0.06	21.71±0.04	20.67±0.06	13	1.85	0.51	0.29	0.000	-0.755
TS	18	15	11	7.76	0.73	0.52	14.99±0.01	14.98±0.01	18	2.92	0.39	0.52	0.000	0.258
T1	26	23	15	8.97	0.65	0.39	18.37±0.04	14.58±0.06	14	2.10	0.21	0.24	0.272	0.138
T3	32	32	31	30.12	0.97	0.94	29.63±0.04	28.52±0.06	14	2.13	0.45	0.47	0.335	0.032

CR/n and CD/n refers to clonal richness and diversity, respectively, corrected for sample size expressed as proportion of clones to total individuals genotyped. A = number of alleles; Ar = allelic richness; H_o = observed heterozygosity; H_e = expected heterozygosity; F_{IS} = fixation index between individuals within local populations. [¥]The numbers are the p-value from a goodness of fit to HWE expectations test using Fisher's exact test method. [$]refers to the expected clonal richness (CR) and clonal diversity (CD) under Equilibrium using randomisation tests implemented in Hwclon (De Meester & Vanoverbeke 1999)There is no significant difference (at α= 0.05) between Observed CR/CD and expected CR/CD values for all population comparisons

Table 3: Summary results of Colonize analyses with three prior allele frequency distributions (Freq, Rich, Level; see main text for explanation), showing the most likely number of founders for each population, based on either mtDNA or nDNA data, for each population and averaged over all populations. Sd: standard deviation. Values with asterisk indicate that the likelihood score was too low ($p < 0.05$) to represent a reliable estimate. Non-integer values represent the average of shared highest scores.

	N° of founders with highest likelihood score (Colonize)					
	mtDNA			nDNA		
Water body	Freq	Rich	Level	Freq	Rich	Level
Adi Gela	4*	4	4	2*	2	2
Adi Kenafiz	5*	8	7	7*	7	5
Dibla	5	4	4	5	3	2
Gereb Awso	1	1	1	5	2	2
Gereb Mihiz	5	6	4	12.5	4	4
Gum Selasa	8	8	7.5	13	5	4
Haiba	2.5	2	2	3*	2	2
Mai Leba	4	3.5	3	5	3	2
Meala	2	2	2	2*	2	2
Tsinkanet	1	1	1	3*	3*	4
T1	5	5	5	3*	2	2
T3	1	1	1	6	3	2
Average	3.63	3.79	3.46	5.54	3.17	2.75
Standard deviation	2.17	2.55	2.23	3.71	1.53	1.14

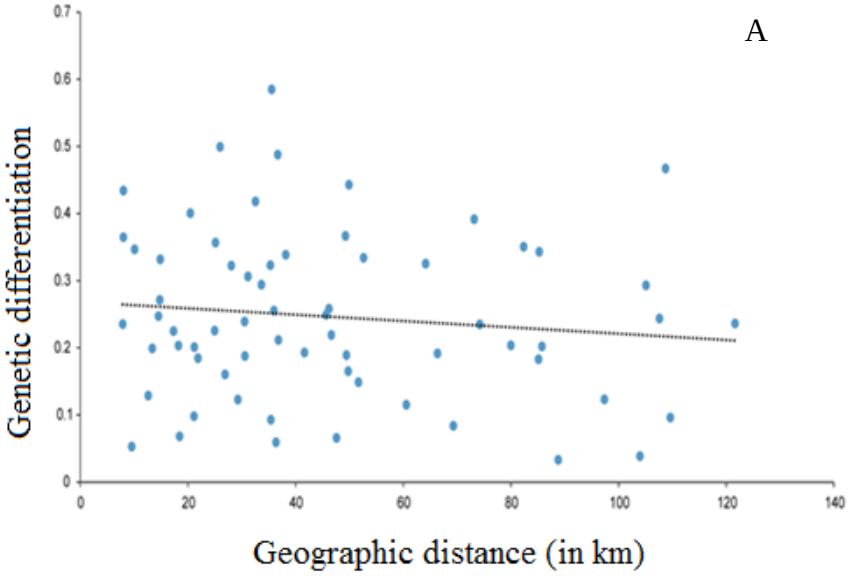


672

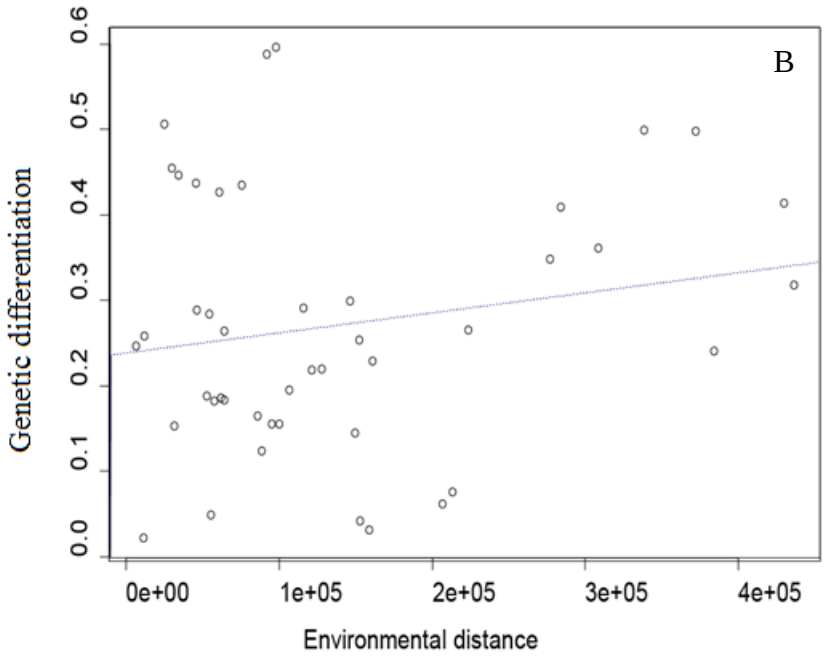
673

674 Figure 1: Geographic location of the sampling sites and mtDNA haplotype frequencies in each
675 population. Major cities are indicated with a star. Inset on the right shows the overall regional
676 frequency of the six encountered haplotypes. AG = Adi Gela; AK = Adi Kenafiz; Di = Dibla; GA
677 = Gereb Awso; GM = Gereb Mihiz; Ha = Haiba; ML = Mai Leba; Me = Meala; Ts = Tsinkanet;
678 T1 = Temporary pond 1; T3 = Temporary pond 3.

679



680



681

682 Figure 2. Relationship between Nei's genetic distance and geographic distance (panel A; testing
683 for an isolation-by-distance and thus for dispersal limitation; r = -0.114; p = 0.662) and the
684 Euclidean distance for environmental variables (panel B; testing for isolation-by-environment; r =
685 0.181; p= 0.212).

686